A non-convex perspective on calibration and imaging in radio interferometry

Proceedings Volume 10394, Wavelets and Sparsity XVII; 103941W (2017)
https://doi.org/10.1117/12.2273766

A non-convex perspective on calibration and imaging in radio interferometry

Audrey Repetti and Yves Wiaux

ISSS, Heriot-Watt University, Edinburgh EH14 4AS, UK

ABSTRACT

New generations of imaging devices aim to produce high resolution and high dynamic range images. In this context, the associated high dimensional inverse problems can become extremely challenging from an algorithmic view point. Moreover, the imaging procedure can be affected by unknown calibration kernels. This leads to the need of performing joint image reconstruction and calibration, and thus of solving non-convex blind deconvolution problems. In this work, we focus on the case where the observed object is affected by smooth calibration kernels in the context of radio astronomy, and we leverage a block-coordinate forward-backward algorithm, specifically designed to minimize non-smooth non-convex and high dimensional objective functions.

Keywords: Non-convex optimization, inverse problem, blind deconvolution, radio interferometric imaging.

1. INTRODUCTION

In the context of image processing, the objective is to find an estimation of an original unknown image $\overline{\boldsymbol{x}} \in \mathbb{R}^N$ from degraded observations $\boldsymbol{y} \in \mathbb{C}^M$ obtained from an imaging device. More formally, these observations can be formulated as an inverse problem, given by

$$\boldsymbol{y} = \overline{\mathcal{A}}(\overline{\boldsymbol{x}}) + \boldsymbol{w}, \tag{1}$$

where $\overline{\mathcal{A}}\colon \mathbb{R}^N \to \mathbb{C}^M$ is the measurement operator and $\boldsymbol{w} \in \mathbb{C}^M$ is a realization of an additive noise. In the theoretical case when $\overline{\mathcal{A}}$ is perfectly known, problem (1) can be solved efficiently using convex optimization techniques.[1] Indeed, a common approach to solve (1) consists in defining the estimate $\boldsymbol{x}^\star$ of $\overline{\boldsymbol{x}}$ as a minimizer of an objective function corresponding to the sum of two terms: the data fidelity term related to the data model, and the regularization term incorporating prior information on the target image. During the last decade, several proximal algorithms have been developed to tackle this minimization problem in a large dimensional context. Among them, we can cite e.g. the forward-backward algorithm (also known as proximal gradient method, or ISTA)[2–4] and primal-dual approaches.[5–8] Nevertheless, in many image processing areas (e.g. astronomical imaging,[9,10] medical imaging,[11] etc.) the measurement operator $\overline{\mathcal{A}}$ is only partially known or completely unknown, leading to the need for calibration methods.

In this work, we focus on the astronomical radio-interferometry (RI) calibration and imaging problem. A radio-interferometer consists of an array of n_a antennas, measuring the radio emission from a given area of the sky. More precisely, the complex noisy measurements $\boldsymbol{y}$ are related to an undersampled selection of the Fourier coefficients of the intensity image $\overline{\boldsymbol{x}}$ degraded by antenna gains. These gains, related to each antenna of the interferometer, are classified as direction-independent and dependent effects (DIEs and DDEs, respectively). On the one hand, DDEs are time-variable complex gains, different for each antenna and varying across the field of view. They correspond to a multiplication in the image domain, and hence a convolution in the Fourier domain. On the other hand, DIEs stand for a particular case of the DDEs, when the spatial dependency can be ignored at each instant of the integration (i.e. only a scalar complex gain factor is considered for each antenna). Note however that traditionally, the calibration process focusses on calibration of DIEs only.

When the antenna gains are assumed to have been pre-calibrated, performing calibration transfer to obtain normalized DIEs, the imaging problem relies on solving an inverse problem of the form of (1), where $\overline{\mathcal{A}}$ is

Further author information: (Send correspondence to A.R.)
A.R.: E-mail: a.repetti@hw.ac.uk
Y.W.: E-mail: y.wiaux@hw.ac.uk

assumed to be approximately known . To this aim, several efficient methods have been developed since the early 1970s, starting with the celebrated CLEAN algorithm.[12–14] However, these pioneering methods are no longer adapted to the next generation of radio telescopes, such as LOw Frequency ARray (LOFAR)*, Meer Karoo Array Telescope (MeerKAT)†, and the future Square Kilometre Array (SKA)†, which are envisaged to produce giga-pixel images and achieve a dynamic range of six or seven orders of magnitude. On the one hand, the number of new methods has significantly increased the last decade to scale the large dimensional problem involved in RI. Mainly these approaches are either based on variants of CLEAN algorithm (e.g. Multi-Scale CLEAN[15] and Adaptive Scale Pixel-CLEAN[16]), or leverage compressive sensing and optimization theories[17–23] by exploiting the fact that many images of the sky are sparse in some dictionary (e.g. image domain, possibly redundant wavelet basis,[24] gradient domain,[25,26] etc.). On the other hand, to achieve higher dynamic range and to match the imaging capabilities of the new generation radio telescopes, estimating only the DIEs is no longer sufficient and the DDEs must be incorporated in the calibration process. Therefore, new calibration techniques accounting for the DDEs have been developed recently. In particular, the Source Peeling and Atmospheric Modeling (SPAM) method has been proposed[27] to iteratively solve and correct for ionospheric phase errors. Basically, SPAM solves for complex gains towards a number of bright sources, then approximates the DDEs by a sum of low-order Zernike polynomials, fitting the phase component of the solutions. Furthermore, traditional DIE calibration methods have been improved by the SAGE algorithm,[28,29] while another general framework has been developed solving for the associated non-linear Least Squares (LS) problem,[30] considering the complex Jacobian formalism. In addition, a different approach based on non-linear Kalman filter has been developed.[31] Finally, a new calibration scheme based on facet calibration, has been developed in the last years.[30,32,33] To apply this method, it is assumed that the DDEs are piece-wise constant in the image domain. Therefore, the sky is partitioned into facets, with the facet center determined by the brightest source or the approximate center of a source group. At each step, the calibration solutions are obtained for the facet center, which is then applied over the whole facet. It is worth mentioning that, to the best of our knowledge, this method has only been applied to point sources models of the sky, which at most include few extended sources. Also note that, when this DDE calibration method is combined with an imaging method, the obtained global reconstruction algorithm does not benefit from any global convergence guarantee.

In our recent work[10] we proposed a new RI method to estimate jointly the image and the DDEs. Our approach is inspired by both the imaging techniques using optimization and compressive sensing theories. It aims to minimize a regularized non-linear LS criterion, with respect to both the image and the DDEs. Note that the DDE calibration part can be seen as a generalization of the alternating DIE calibration method StEFCal.[34,35] On the one hand, concerning the DDEs, we assume that they are smooth functions across the field of view, i.e. spatially band-limited. Therefore we propose to estimate only the non-zero Fourier coefficients of the DDEs within a known fixed support. Then, this assumption not only provides a strong regularization for the estimation of the DDEs but also allows to reduce drastically the dimension of the problem. On the other hand, concerning the intensity image, we use a hybrid regularization term relying on a positivity constraint and an ℓ_1 regularization term to promote the sparsity of the image (or its sparsity in a given dictionary). Therefore, an algorithm has been designed based on recent non-convex optimization techniques.[36–38] It consists in alternating between the estimation of the image and the DDEs, using forward-backward iterations (i.e. a gradient step followed by a proximity step). One of the main advantage of our method is that it benefits from convergence guarantees, in particular that the sequence of iterates generated by the algorithm converges to a critical point of the global objective function. Moreover it works globally on the whole image and in an automatic manner, without any requirement of partitioning the sky into different directions. Furthermore, it is adapted not only to point sources images, but also to reconstruct images with extended and complex structures. Finally, the results of our preliminary simulations presented in[10,39,40] suggest that using our method to jointly estimate the DDEs and the image leads to large improvement of the dynamic range, that is by orders of magnitude compared to accounting for DIEs only.

Nevertheless, in our previous works,[10,39] we made use of the assumption that the brightest sources of the image are known exactly, or at most can be corrupted by small variations in the amplitude in the context of

*http://www.lofar.org/

†http://www.skatelescope.org/

images with point sources.[40] This assumption was useful to reduce the ambiguity problems. In particular, the joint calibration and imaging problem being a blind deconvolution problem, any arbitrary modulation of the intensity image can be absorbed in the DDEs without affecting the data in equation (1). The data are also blind to a common translation of the DDE kernels in the Fourier domain. These two points have been mitigated by the assumption that both the intensity and the position of the brightest sources in the sky was assumed to be known. However in practice it is difficult to gather such information and the objective of this work is to relax it when performing joint DDE calibration and imaging reconstruction. In particular, we consider a realistic setting where our prior information consists of approximated values of the zero spatial frequency coefficients of the DDEs (i.e. the DIEs) obtained by performing calibration transfer. This process consists in performing DIE calibration on a known source out of the field of view of interest, between two acquisition time instants. Then, the estimated DIEs are interpolated to obtain an approximation of the DIEs appearing in the data of interest. Finally, normalized DIEs are obtained by dividing each acquisition by the approximated DIEs, and they can be used as prior information. In this work, we show through simulation results that our joint DDE calibration and imaging method has good performance in this context, considering the MeerKAT antenna distribution and sophisticated images with complex structures.

The remainder of the paper is organized as follows. The RI problem is described in Section 2. In Section 3 we give the associated minimization problem and a description of the proposed optimization method. To show the good behaviour of our approach, simulations and results are presented in Section 4. Finally, some conclusions are draxn in Section 5.

2. PROBLEM STATEMENT

2.1 Radio interferometric data

In this section we describe the observation model, accounting for the DDEs, in the RI problem. The complex observations, namely the visibilities, acquired at instant $t \in \{1, \dots, T\}$, are determined by the relative position between each antenna pair indexed by $(\alpha, \beta) \in \{1, \dots, n_a\}^2$, with $\alpha < \beta$. Therefore, an interferometer acquires $M = Tn_a(n_a - 1)/2$ different measurements in the Fourier domain of the image of interest. Moreover, each visibility $y_{t,\alpha,\beta} \in \mathbb{C}$ measured by the antenna pair (α, β) at instant t at the discrete spatial frequency $k_{t,\alpha,\beta}$ can be modelled as

$$y_{t,\alpha,\beta} = \sum_{n=-N/2}^{N/2-1} \overline{d}_{t,\alpha}(n)\overline{d}_{t,\beta}(n)^* \overline{x}(n) e^{-2\mathrm{i}\pi k_{t,\alpha,\beta}\frac{n}{N}} + w_{t,\alpha,\beta}, \tag{2}$$

where $\overline{\boldsymbol{d}}_{t,\alpha} = (\overline{d}_{t,\alpha}(n))_{-N/2 \leq n \leq N/2-1} \in \mathbb{C}^N$ represents the DDE related to antenna α, and $\boldsymbol{w} = (w_{t,\alpha,\beta})_{\substack{1 \leq t \leq T \\ 1 \leq \alpha < \beta \leq n_a}} \in \mathbb{C}^M$ is a realization of a complex i.i.d. Gaussian additive noise. Note that, using these notations, the DIEs can be seen as a special case of the DDEs where $\overline{\boldsymbol{d}}_{t,\alpha} = \delta_{t,\alpha}\mathbf{1}_N$, with $\delta_{t,\alpha} \in \mathbb{C}$ and $\mathbf{1}_N$ being the unitary vector of dimension N.

2.2 Prior information

In this work we focus on the case when the DDEs correspond to smooth functions across the field of view, i.e. they are spatially band-limited. In other words, we assume that, for every $(\alpha, t) \in \{1, \dots, n_a\} \times \{1, \dots, T\}$, the 2D Fourier transform $\widehat{\overline{\boldsymbol{d}}}_{t,\alpha}$ of the DDE $\overline{\boldsymbol{d}}_{t,\alpha}$, associated with antenna α at time instant t, has a bounded known support of dimension $S \ll N$. Then, $\overline{\boldsymbol{d}}_{t,\alpha}$ is characterized by its non-zero Fourier coefficients denoted by $\overline{\boldsymbol{u}}_{t,\alpha} \in \mathbb{C}^S$. Moreover, in order to simplify notation in the remainder of the paper, let us introduce, for every $t \in \{1, \dots, T\}$, the matrix $\overline{\boldsymbol{U}}_t = \left(\overline{\boldsymbol{u}}_{t,\alpha}\right)_{1 \leq \alpha \leq n_a} \in \mathbb{C}^{n_a \times S}$, representing the concatenation of the non-zero Fourier coefficients of the DDEs.

As described in the introduction, we assume that calibration transfer has been performed, and thus that the DIEs are normalized. In other words, for every $t \in \{1, \dots, T\}$ and $\alpha \in \{1, \dots, n_a\}$, we can assume that the zero spatial frequency coefficient $\overline{u}_{t,\alpha}(0)$ (i.e. the DIE) belongs to a complex neighbourhood of $1 + \mathrm{i}0$. Moreover, we assume that the other coefficients of $\overline{\boldsymbol{u}}_{t,\alpha}$ belong to a complex neighbourhood of 0. Note that it is reasonable to

consider small values for the higher order spatial frequencies since they represent direction-dependent variations in the gain across the field of view with respect to the mean gain.

It is important to emphasize that having normalized DIEs can also be used to obtain prior information on the image. Indeed, even though the forward model described in (2) take into account DDEs, we can approximate it by considering that, for every $t \in \{1,\dots,T\}$ and $\alpha \in \{1,\dots,n_a\}$, $\delta_{t,\alpha} = 1$. Therefore, an approximation of the estimate of $\overline{\boldsymbol{x}}$ can be obtained by defining it as a solution to

$$\underset{\boldsymbol{x}\in\mathbb{R}^N}{\text{minimize}} \ \frac{1}{2}\|\mathcal{F}(\boldsymbol{x}) - \boldsymbol{y}\|_2^2 + \widetilde{r}(\boldsymbol{x}), \tag{3}$$

where $\mathcal{F}\colon \mathbb{C}^N \to \mathbb{C}^K$ represents the 2D discrete Fourier transform operator, and $\widetilde{r}\colon \mathbb{R}^N \to]-\infty,+\infty]$ is a regularization function. Let $\boldsymbol{x}_0^\star$ be a minimizer of problem (3). It is interesting to note that $\boldsymbol{x}_0^\star$ is the best estimation which can be obtained (associated with the regularization term $\widetilde{r}$) if DDE calibration is not performed. Then, $\boldsymbol{x}_0^\star$ can be used as prior information on the image to perform joint DDE calibration and imaging. Nevertheless, due to the fact that DDEs are not taken into account in (3), $\boldsymbol{x}_0^\star$ will contain artefacts (e.g. noisy background, wrong amplitude, wrong source detection, etc.). Therefore, in our approach, we propose to use a thresholded version of $\boldsymbol{x}_0^\star$, denoted by $\boldsymbol{x}_0$, where low amplitude coefficients have been removed. Then, the original unknown image can be decomposed as the sum of $\boldsymbol{x}_0$ and another image, denoted by $\overline{\boldsymbol{\epsilon}} \in \mathbb{R}^N$ which has to be estimated, i.e. $\overline{\boldsymbol{x}} = \boldsymbol{x}_0 + \overline{\boldsymbol{\epsilon}}$. Thus, problem (2) can be rewritten as follows:

$$y_{t,\alpha,\beta} = \sum_{n=-N/2}^{N/2-1} \overline{d}_{t,\alpha}(n)\overline{d}_{t,\beta}(n)^*\big(x_0(n) + \overline{\epsilon}(n)\big)e^{-2\mathrm{i}\pi k_{t,\alpha,\beta}\frac{n}{N}} + w_{t,\alpha,\beta}. \tag{4}$$

2.3 Inverse problem

The inverse problem described by equation (4) is non linear in $\Big(\overline{\boldsymbol{\epsilon}}, (\overline{\boldsymbol{d}}_{\alpha,t})_{\substack{1\le t\le T\\ 1\le\alpha<\beta\le n_a}}\Big)$. However, it is linear in $\overline{\boldsymbol{\epsilon}}$, and (2) can be rewritten equivalently as follows:

$$\boldsymbol{y} = \overline{\boldsymbol{G}}\mathcal{F}(\boldsymbol{x}_0 + \overline{\boldsymbol{\epsilon}}) + \boldsymbol{w}, \tag{5}$$

where $\overline{\boldsymbol{G}} \in \mathbb{C}^{M\times K}$ is the matrix containing on each row the antenna-based gain for the pair (α,β). More precisely, each row of $\overline{\boldsymbol{G}}$, indexed by $\big(t,\alpha,\beta\big)$, corresponds to the convolution of the Fourier transforms $\widehat{\overline{\boldsymbol{d}}}_{t,\alpha}$ and $\widehat{\overline{\boldsymbol{d}}^*}_{t,\beta}$ of the gains $\overline{\boldsymbol{d}}_{t,\alpha}$ and $\overline{\boldsymbol{d}}^*_{t,\beta}$ respectively, centred at the spatial frequency $k_{t,\alpha,\beta}$. Note that problem (5) can be solved efficiently when the antenna based gains are assumed to be known. However, in practice, only a non-accurate approximation of $\overline{\boldsymbol{G}}$ is accessible, and the objective of this work is to find a better estimate of $\overline{\boldsymbol{G}}$ along with an estimate of $\overline{\boldsymbol{\epsilon}}$. To this aim, we adopt the same strategy as described in our previous work,[10] which is a generalization of the StEFCal model.[35] In particular, we propose to rewrite the problem described by (2) using a matrix formulation:

$$(\forall t \in \{1,\dots,T\}) \quad \boldsymbol{Y}_t = \mathcal{D}_{1,t}\big(\overline{\boldsymbol{U}}_t\big)\mathcal{X}\big(\mathcal{F}(\boldsymbol{x}_0 + \overline{\boldsymbol{\epsilon}})\big)\mathcal{D}_{2,t}\big(\overline{\boldsymbol{U}}_t\big) + \boldsymbol{W}_t, \tag{6}$$

where, for every $t \in \{1,\dots,T\}$, $\mathcal{D}_{1,t}\colon \mathbb{C}^{n_a\times S} \to \mathbb{C}^{n_a\times N}$ (resp. $\mathcal{D}_{2,t}\colon \mathbb{C}^{n_a\times S} \to \mathbb{C}^{n_a\times N}$) is the operator defined such that $\mathcal{D}_{1,t}\big(\overline{\boldsymbol{U}}_t\big)$ (resp. $\mathcal{D}_{2,t}\big(\overline{\boldsymbol{U}}_t\big)$) is the sparse matrix containing on each row $\alpha \in \{1,\dots,n_a\}$ the Fourier kernel $\overline{\boldsymbol{u}}_{t,\alpha}$ flipped and centred in k_α (resp. $\overline{\boldsymbol{u}}^*_{t,\alpha}$ centred in $-k_\alpha$), $\mathcal{X}\big(\mathcal{F}(\boldsymbol{x}_0 + \overline{\boldsymbol{\epsilon}})\big) \in \mathbb{C}^{K\times K}$ contains on each row/column a shifted version of the Fourier transform of the original image $\mathcal{F}(\boldsymbol{x}_0 + \overline{\boldsymbol{\epsilon}})$, and $\boldsymbol{W}_t \in \mathbb{C}^{n_a\times n_a}$ is the matrix formulation associated with the realization of the additive noise in equation (4).

Note that, though problem (6) is not linear with respect to the DDEs (i.e. $(\overline{\boldsymbol{U}}_t)_{1\le t\le T}$), we propose to adopt a bilinear approach by introducing, for every $t \in \{1,\dots,T\}$, the matrices $\overline{\boldsymbol{U}}_{1,t} = \big(\overline{\boldsymbol{u}}_{1,t,\alpha}\big)_{1\le\alpha\le n_a} \in \mathbb{C}^{n_a\times S}$ and $\overline{\boldsymbol{U}}_{2,t} = \big(\overline{\boldsymbol{u}}_{2,t,\alpha}\big)_{1\le\alpha\le n_a} \in \mathbb{C}^{n_a\times S}$, satisfying $\overline{\boldsymbol{U}}_{1,t} = \overline{\boldsymbol{U}}_{2,t} = \overline{\boldsymbol{U}}_t$. Then, the objective is to find an estimation of $\Big(\overline{\boldsymbol{\epsilon}}, \overline{\boldsymbol{U}}_1, \overline{\boldsymbol{U}}_2\Big)$, where $\overline{\boldsymbol{U}}_1 = \big(\overline{\boldsymbol{U}}_{1,t}\big)_{1\le t\le T} \in \mathbb{C}^{Tn_a\times S}$ and $\overline{\boldsymbol{U}}_2 = \big(\overline{\boldsymbol{U}}_{2,t}\big)_{1\le t\le T} \in \mathbb{C}^{Tn_a\times S}$.

3. PROPOSED OPTIMIZATION APPROACH

3.1 Global minimization problem

We propose to define the estimate $(\boldsymbol{\epsilon}^\star, \boldsymbol{U}_1^\star, \boldsymbol{U}_2^\star)$ of $(\overline{\boldsymbol{\epsilon}}, \overline{\boldsymbol{U}}_1, \overline{\boldsymbol{U}}_2)$ as a solution to the following non-convex minimization problem:

$$\underset{\boldsymbol{\epsilon}\in\mathbb{R}^N,\,(\boldsymbol{U}_1,\boldsymbol{U}_2)\in(\mathbb{C}^{n_a\times S})^{2T}}{\text{minimize}} \; h\big(\boldsymbol{\epsilon},\boldsymbol{U}_1,\boldsymbol{U}_2\big) + r(\boldsymbol{\epsilon}) + p\big(\boldsymbol{U}_1,\boldsymbol{U}_2\big), \tag{7}$$

where $h\colon \mathbb{R}^N \times (\mathbb{C}^{n_a\times S})^{2T} \to \mathbb{R}$ is the data fidelity term related to the model described by equations (5)-(6), $r\colon \mathbb{R}^N \to]-\infty,+\infty]$ is the regularization term associated with the image, and $p\colon (\mathbb{C}^{n_a\times S})^{2T} \to]-\infty,+\infty]$ is the regularization term associated with the DDEs. In particular, the RI measurements being corrupted by an i.i.d. Gaussian additive noise, we choose h to be the least squares criterion associated with equations (5)-(6), i.e.

$$h\big(\boldsymbol{\epsilon},\boldsymbol{U}_1,\boldsymbol{U}_2\big) = \Big\|\mathcal{G}\big(\boldsymbol{U}_1,\boldsymbol{U}_2\big)\mathcal{F}(\boldsymbol{x}_0+\boldsymbol{\epsilon}) - \boldsymbol{y}\Big\|_2^2 \tag{8}$$

$$= \frac{1}{2}\sum_{t=1}^{T}\Big\|\mathcal{D}_{1,t}\big(\boldsymbol{U}_{1,t}\big)\mathcal{X}\big(\mathcal{F}(\boldsymbol{x}_0+\boldsymbol{\epsilon})\big)\mathcal{D}_{2,t}\big(\boldsymbol{U}_{2,t}\big) - \boldsymbol{Y}_t\Big\|, \tag{9}$$

with $\mathcal{G}\colon (\mathbb{C}^{n_a\times S})^{2T} \to \mathbb{C}^{M\times K}$ being the operator building a matrix containing on each row, indexed by (t,α,β), the convolution of the Fourier kernels $\boldsymbol{u}_{1,t,\alpha}$ and $\boldsymbol{u}^*_{2,t,\beta}$ centred at the spatial frequency $k_{t,\alpha,\beta}$. In particular we have $\mathcal{G}\big(\overline{\boldsymbol{U}}_1,\overline{\boldsymbol{U}}_2\big) = \mathcal{G}\big(\overline{\boldsymbol{U}},\overline{\boldsymbol{U}}\big) = \overline{\boldsymbol{G}}$, where $\overline{\boldsymbol{G}}$ is the matrix defined in equation (5).

We propose to use a hybrid function for the regularization term of the image, corresponding to the sum of a sparsity regularization term and an indicator function introducing constraints on the coefficients of the image. More formally, we choose

$$(\forall \boldsymbol{\epsilon}\in\mathbb{R}^N) \quad r(\boldsymbol{\epsilon}) = \lambda\|\boldsymbol{\Psi}^\dagger(\boldsymbol{x}_0+\boldsymbol{\epsilon})\|_1 + \iota_{\mathbb{E}}(\boldsymbol{\epsilon}), \tag{10}$$

where $\boldsymbol{\Psi}\in\mathbb{R}^{Q\times N}$ is a given sparsity basis, $\lambda>0$ is a regularization parameter, $\mathbb{E}$ is a closed, convex and non-empty subset of $\mathbb{R}^N$, and $\iota_{\mathbb{E}}$ denotes the indicator function of set $\mathbb{E}$ defined as

$$(\forall \boldsymbol{\epsilon}\in\mathbb{R}^N) \quad \iota_{\mathbb{E}}(\boldsymbol{\epsilon}) = \begin{cases} 0, & \text{if } \boldsymbol{\epsilon}\in\mathbb{E}, \\ +\infty, & \text{otherwise.} \end{cases} \tag{11}$$

Different dictionaries can be chosen, depending on the nature of the target image. On the one hand, if the image is assumed to have only few non-zero coefficients (e.g. point sources), one can choose $\boldsymbol{\Psi} = \boldsymbol{I}_N$ to be the identity operator in $\mathbb{R}^N$. On the other hand, natural images are not necessarily sparse, but can have a sparse representation in other domains. For instance, piece-wise constant images are sparse on the gradient domain, and thus total variation based regularizations can be used for such images.[25,26,41] Another choice for $\boldsymbol{\Psi}$ is to promote sparsity in the wavelet domain, using, e.g., isotropic undecimated wavelet (IUW) transforms,[20,22,42,43] or a concatenation of wavelet transforms.[21,44] In our case, we propose to use the average sparsity basis[21] which is the concatenation of a Dirac basis with the first eight Deaubechies wavelets.[24] Furthermore, the constraint set $\mathbb{E}$ is chosen to impose positivity on the coefficients of $\boldsymbol{\epsilon}$, and to make use of the prior information $\boldsymbol{x}_0$ we have on $\overline{\boldsymbol{x}}$. Indeed, the image $\boldsymbol{x}_0$ obtained by the process described in Section 2.2 contains information on the brightest sources of $\overline{\boldsymbol{x}}$. Although $\boldsymbol{x}_0$ already gives a good approximation of $\overline{\boldsymbol{x}}$, we propose to choose $\mathbb{E}$ such that $\boldsymbol{\epsilon}$ can be used to absorb the errors appearing on the non-zero coefficients of $\boldsymbol{x}_0$. Let $\mathbb{S}_0$ be the support of $\boldsymbol{x}_0$ and $\mathbb{S}_0^c$ its complementary set. Then, we choose

$$\mathbb{E} = \Big\{\boldsymbol{\epsilon}\in\mathbb{R}^N \,\Big|\, (\forall n\in\mathbb{S}_0)\; -\vartheta x_0(n) \le \epsilon(n) \le \vartheta x_0(n), \text{ and } (\forall n\in\mathbb{S}_0^c)\; 0\le\epsilon(n)\Big\}, \tag{12}$$

where $\vartheta\in[0,1]$ represents the percentage error we assume on $\boldsymbol{x}_0$. In particular, choosing $\vartheta=1$ allows $\epsilon(n)$ to be 0 even when $n\in\mathbb{S}$ (which is not the case when $\vartheta<1$). Choosing $\vartheta=0$ is equivalent to impose that $\boldsymbol{\epsilon}$ is equal to $\boldsymbol{x}_0$ on $\mathbb{S}$. This last choice corresponds to the particular case investigated in[10] when bright sources are assumed to be exactly known.

Concerning the regularization term for the DDEs, we choose it to constrain the coefficients of $(\boldsymbol{U}_1,\boldsymbol{U}_2)$ to satisfy the assumption given in Section 2.2, and we want to oblige $\boldsymbol{U}_1$ and $\boldsymbol{U}_2$ to be equal at convergence. For

the first constraint, we define $\mathbb{D}$ to be the subset of $\mathbb{C}^N$ such that, for every $t \in \{1, \dots, T\}$ and $\alpha \in \{1, \dots, n_a\}$, $u_{1,t,\alpha}(0)$ and $u_{2,t,\alpha}(0)$ belong to an ℓ_∞ complex ball centred in $1 + \mathrm{i}0$ with radius υ, while, for every $s \neq 0$, $u_{1,t,\alpha}(s)$ and $u_{2,t,\alpha}(s)$ belong to an ℓ_∞ complex ball centred in 0 with radius υ. Moreover, we want to impose that the Fourier coefficients of the DDEs have a mean $\boldsymbol{\Theta} \in \mathbb{C}^{Tn_a \times S}$ which is defined such that each line of this matrix is equal to $1 + \mathrm{i}0$ for the central spatial frequencies, and 0 for the higher frequencies. To this aim, we use an ℓ_2 norm controlling the distance between $\boldsymbol{U}_1$ (resp. $\boldsymbol{U}_2$) and $\boldsymbol{\Theta}$. Finally, concerning the constraint that we want to impose $\boldsymbol{U}_1$ and $\boldsymbol{U}_2$ to be equal, due to technical assumptions on the algorithm we have developed, we propose to relax this constraint by choosing to minimize a squared distance between these two variables. Therefore, the global regularization term chosen for the DDEs is given by

$$\big(\forall (\boldsymbol{U}_1, \boldsymbol{U}_2) \in (\mathbb{C}^{n_a \times S})^{2T}\big) \quad p\big(\boldsymbol{U}_1, \boldsymbol{U}_2\big) = \eta \|\boldsymbol{U}_1 - \boldsymbol{U}_2\|_2^2 + \mu\Big(\|\boldsymbol{U}_1 - \boldsymbol{\Theta}\|_2^2 + \|\boldsymbol{U}_2 - \boldsymbol{\Theta}\|_2^2\Big) + \iota_{\mathbb{D}}\big(\boldsymbol{U}_1\big) + \iota_{\mathbb{D}}\big(\boldsymbol{U}_2\big), \quad (13)$$

where $\mu > 0$ and $\eta > 0$ are regularization parameters. Note that the smooth regularization can be seen as reducing the degeneracy relative to the fact the a modulation of the image can be absorbed in the DDEs without affecting the data.

3.2 Optimization tools

In order to describe our optimization method to solve the joint calibration and imaging problem given in equation (7), we first present in this section the basics of optimization. Note that the definitions given below stand for non-necessarily convex functions. We refer the reader to[45,46] for an overview in convex optimization, and to[47,48] for non-convex optimization.

A function $\psi \colon \mathbb{R}^Q \to]-\infty, +\infty]$ is proper if its domain, denoted by $\operatorname{dom} \psi$, is non-empty.

Let ψ be differentiable on $\mathbb{R}^Q$. The function ψ is Lipschitz-differentiable, with constant $\zeta > 0$, if

$$(\forall (\boldsymbol{u}, \boldsymbol{z}) \in \mathbb{R}^Q) \quad \|\nabla \psi(\boldsymbol{u}) - \nabla \psi(\boldsymbol{z})\|_2 \leq \zeta \|\boldsymbol{u} - \boldsymbol{z}\|_2. \quad (14)$$

Let ψ be a proper, lower-semicontinuous function, bounded from below by an affine function. Its proximity operator[49] defined at $\boldsymbol{z} \in \mathbb{R}^Q$, denoted by $\operatorname{prox}_\psi(\boldsymbol{z})$, corresponds to the set of minimizers of $\psi + \frac{1}{2}\|\cdot - \boldsymbol{z}\|_2^2$, i.e.

$$\operatorname{prox}_\psi(\boldsymbol{z}) = \underset{\boldsymbol{u} \in \mathbb{R}^Q}{\operatorname{argmin}} \ \psi(\boldsymbol{u}) + \frac{1}{2}\|\boldsymbol{u} - \boldsymbol{z}\|_2^2. \quad (15)$$

Note that, in the case when ψ is convex, $\operatorname{prox}_\psi(\boldsymbol{z})$ reduces to a singleton. Proximity operators are used in optimization algorithms to deal with non-smooth functions, whereas differentiable functions are usually handled by computing gradient steps. A particular case of proximity operator is the projection operator onto a closed non-empty set $\mathbb{E}$, denoted by $\Pi_{\mathbb{E}}$, obtained when ψ is the indicator function of $\mathbb{E}$.

3.3 Proposed algorithm

To solve problem (7), one can exploit the block-variable structure of the objective function by using an alternating forward-backward algorithm.[36,38,50–52] Basically, this algorithm alternates between the estimation of the unknown image $\overline{\boldsymbol{\epsilon}}$ and the estimation of the DDEs represented by the matrices $\overline{\boldsymbol{U}}_1$ and $\overline{\boldsymbol{U}}_2$. At each iteration, this approach combines a gradient step (forward step) on the Lipschitz-differentiable functions with a proximity step (backward step) on the non-smooth functions.

In the context of problem (7), on the one hand, for the image estimation, at each iteration, a gradient step is performed on the differentiable data fidelity term h while a proximity step is computed to deal with the non-smooth regularization function r. On the other hand, for the DDEs estimation, at each iteration $i \in \mathbb{N}$, the forward step involves not only the data fidelity term but also the differentiable part of the function p, i.e. the squared distance between $\boldsymbol{U}_1^{(i)}$ and $\boldsymbol{U}_2^{(i)}$ and between $\boldsymbol{U}_1^{(i)}$ (resp. $\boldsymbol{U}_2^{(i)}$) and $\boldsymbol{\Theta}$. Then a projection step onto set $\mathbb{D}$ is performed.

The proposed method is described in Algorithm (1), where $\cdot$ denotes the Hadamard product, and, for every $i \in \mathbb{N}$, $\tau^{(i)}$ is a positive constant, and $\Gamma_1^{(i)}$ and $\Gamma_2^{(i)}$ are positive matrices.

Algorithm 1 Alternating forward-backward algorithm

Initialization Let $\boldsymbol{\epsilon}^{(0)} \in \operatorname{dom} r$ and $(\boldsymbol{U}_1^{(0)}, \boldsymbol{U}_2^{(0)}) \in \mathbb{D}^2$. Let, for every $i \in \mathbb{N}$, $(L^{(i)}, J^{(i)}) \in \mathbb{N}^2$.

Iterations

For $i = 0, 1, \ldots$

  Choose to update either the DDEs $\big(\boldsymbol{U}_1^{(i)}, \boldsymbol{U}_2^{(i)}\big)$, or the image $\boldsymbol{\epsilon}^{(i)}$.

  If the DDEs are updated:

    $\boldsymbol{U}_1^{(i,0)} = \boldsymbol{U}_1^{(i)}$, $\boldsymbol{U}_2^{(i,0)} = \boldsymbol{U}_2^{(i)}$.

    For $\ell = 0, \ldots, L^{(i)} - 1$

      $\boldsymbol{U}_1^{(i,\ell+1)} = \Pi_{\mathbb{D}}\Big(\boldsymbol{U}_1^{(i,\ell)} - \Gamma_1^{(i)} \cdot \Big(\nabla_{\boldsymbol{U}_1} h\big(\boldsymbol{\epsilon}^{(i)}, \boldsymbol{U}_1^{(i,\ell)}, \boldsymbol{U}_2^{(i)}\big) - \eta\big(\boldsymbol{U}_1^{(i,\ell)} - \boldsymbol{U}_2^{(i)}\big) - \mu\big(\boldsymbol{U}_1^{(i,\ell)} - \boldsymbol{\Theta}\big)\Big)\Big)$,

    end for

    $\boldsymbol{U}_1^{(i+1)} = \boldsymbol{U}_1^{(i,L^{(i)})}$.

    For $\ell = 0, \ldots, L^{(i)} - 1$

      $\boldsymbol{U}_2^{(i,\ell+1)} = \Pi_{\mathbb{D}}\Big(\boldsymbol{U}_2^{(i,\ell)} - \Gamma_2^{(i)} \cdot \Big(\nabla_{\boldsymbol{U}_2} h\big(\boldsymbol{\epsilon}^{(i)}, \boldsymbol{U}_1^{(i+1)}, \boldsymbol{U}_2^{(i,\ell)}\big) - \eta\big(\boldsymbol{U}_2^{(i,\ell)} - \boldsymbol{U}_1^{(i+1)}\big) - \mu\big(\boldsymbol{U}_2^{(i,\ell)} - \boldsymbol{\Theta}\big)\Big)\Big)$,

    end for

    $\boldsymbol{U}_2^{(i+1)} = \boldsymbol{U}_2^{(i,L^{(i)})}$.

    $\boldsymbol{\epsilon}^{(i+1)} = \boldsymbol{\epsilon}^{(i)}$.

  end if

  If the image is updated:

    $\boldsymbol{\epsilon}^{(i,0)} = \boldsymbol{\epsilon}^{(i)}$.

    For $j = 0, \ldots, J^{(i)} - 1$

      $\boldsymbol{\epsilon}^{(i,j+1)} = \operatorname{prox}_{\tau^{(i)} r}\Big(\boldsymbol{\epsilon}^{(i,j)} - \tau^{(i)} \nabla_{\boldsymbol{\epsilon}} h\big(\boldsymbol{\epsilon}^{(i,j)}, \boldsymbol{U}_1^{(i)}, \boldsymbol{U}_2^{(i)}\big)\Big)$,

    end for

    $\boldsymbol{\epsilon}^{(i+1)} = \boldsymbol{\epsilon}^{(i,J^{(i)})}$.

    $\big(\boldsymbol{U}_1^{(i+1)}, \boldsymbol{U}_2^{(i+1)}\big) = \big(\boldsymbol{U}_1^{(i)}, \boldsymbol{U}_2^{(i)}\big)$.

  end if

The proposed method is based on a variable metric block-coordinate forward-backward algorithm developed by[38] (other versions can be found in[36,37]). In particular, we can deduce from this paper the convergence guarantees of Algorithm 1 given below.

In order to ensure the convergence of the proposed algorithm, few technical assumptions have to be satisfied. Let $(\boldsymbol{\epsilon}^{(i)}, \boldsymbol{U}^{(i)}, \boldsymbol{U}^{(i)})_{i\in\mathbb{N}}$ be a sequence generated by Algorithm 1. Firstly, we assume that the blocks $(\boldsymbol{\epsilon}^{(i)}, \boldsymbol{U}^{(i)}, \boldsymbol{U}^{(i)})_{i\in\mathbb{N}}$ are updated following an essentially cyclic rule, i.e. each variable has to be updated at least one time within a given finite number of iterations. Secondly, for every $i \in \mathbb{N}$, $\tau^{(i)}$ needs to satisfy

$$0 < \tau^{(i)} < 1/\|\mathcal{G}(\boldsymbol{U}_1^{(i)}, \boldsymbol{U}_2^{(i)}\|_S, \tag{16}$$

where $\|.\|_S$ denotes the spectral norm. Finally, for every $i \in \mathbb{N}$, $\Gamma_1^{(i)}$ and $\Gamma_2^{(i)}$ have to satisfy the majorant condition given by [38, Ass. 2.3]. In particular, in our case we propose to choose

$$\Gamma_1^{(i)} = \begin{bmatrix} \gamma_{1,1,1}^{(i)} \mathbf{1}_S^\top \\ \vdots \\ \gamma_{1,T,n_a}^{(i)} \mathbf{1}_S^\top \end{bmatrix} \quad \text{and} \quad \Gamma_2^{(i)} = \begin{bmatrix} \gamma_{2,1,1}^{(i)} \mathbf{1}_S^\top \\ \vdots \\ \gamma_{2,T,n_a}^{(i)} \mathbf{1}_S^\top \end{bmatrix} \tag{17}$$

where

$$(\forall t \in \{1,\dots,T\})(\forall \alpha \in \{1,\dots,n_a\}) \quad \begin{cases} 0 < \gamma_{1,t,\alpha}^{(i)} < 1/(\eta+\mu+\zeta_{1,t,\alpha}^{(i)}), \\ 0 < \gamma_{2,t,\alpha}^{(i)} < 1/(\eta+\mu+\zeta_{2,t,\alpha}^{(i)}), \end{cases} \tag{18}$$

$\zeta_{1,t,\alpha}^{(i)}$ (resp. $\zeta_{2,t,\alpha}^{(i)}$) being the Lipschitz constant of the partial derivative of $h(\boldsymbol{\epsilon}^{(i)}, \boldsymbol{U}_1^{(i)}, \boldsymbol{U}_2^{(i)})$ (resp. $h(\boldsymbol{\epsilon}^{(i)}, \boldsymbol{U}_1^{(i+1)}, \boldsymbol{U}_2^{(i)})$) with respect to $\boldsymbol{u}_{1,t,\alpha}^{(i)}$ (resp. $\boldsymbol{u}_{2,t,\alpha}^{(i)}$).

Under these assumptions, the sequence of iterates $\left(\boldsymbol{\epsilon}^{(i)}, \boldsymbol{U}_1^{(i)}, \boldsymbol{U}_2^{(i)}\right)_{i\in\mathbb{N}}$ generated by Algorithm 1 converges to a critical point $\left(\boldsymbol{\epsilon}^{\star}, \boldsymbol{U}_1^{\star}, \boldsymbol{U}_2^{\star}\right)$ of the objective function minimized in problem (7). Moreover, the objective function value is decreasing along the iterations.

4. SIMULATIONS AND RESULTS

To show the performance of the proposed method, we computed numerical experiments, implemented in MATLAB[‡]. We consider the $n_a = 64$ antennas of the MeerKAT telescope, where each antenna pair acquires $T = 15$ measurements. Note that Earth rotation is incorporated to track the $u-v$ positions of each resulting baseline by considering a time interval of 6 hours[§]. Moreover, in order to simplify the experiments, we use discrete versions of the associated $u-v$ coverages. This is done by considering the nearest discrete $u-v$ position of each antenna. In our simulations we have considered two images: an image of M31, given in Figure 1 (1a-c), and an image of W28, given in Figure 2 (1a-c). In both cases, (1a) shows the image in linear scale, (1b) the image in log scale, and (1c) is a zoom on a subpart of the image in log scale. Finally, for every $\alpha \in \{1,\dots,n_a\}$, the DDE kernel $\overline{\boldsymbol{u}}_{t,\alpha}$ associated with antenna α at instant $t \in \{1,\dots,T\}$ is simulated randomly in the Fourier domain and we fix the size S of the support of the direction-dependent Fourier coefficients to be equal to 5×5. Note that in our simulations S is assumed to be known.

For both images of M31 and W28 we have performed preliminary simulations considering only the imaging problem in the case when the DDEs are perfectly known. In this case, we propose to define the estimate of $\overline{\boldsymbol{x}}$ as a solution to

$$\underset{\boldsymbol{x}\in\mathbb{R}^N}{\text{minimize}} \;\; \|\overline{\boldsymbol{G}}\mathcal{F}(\boldsymbol{x}) - \boldsymbol{y}\|^2 + \widetilde{r}(\boldsymbol{x}), \tag{19}$$

where $\overline{\boldsymbol{G}}$ and $\boldsymbol{y}$ are defined in equation (5), and $\widetilde{r}\colon \mathbb{R}^N \to]-\infty,+\infty]$ is given by

$$(\forall \boldsymbol{x} \in \mathbb{R}^N) \quad \widetilde{r}(\boldsymbol{x}) = \lambda\|\boldsymbol{\Psi}^\dagger \boldsymbol{x}\|_1 + \iota_{[0,+\infty[^N}(\boldsymbol{x}), \tag{20}$$

with λ and $\boldsymbol{\Psi}$ being defined by (10). Problem (19) can be solved efficiently using a convex optimization method (e.g. forward-backward algorithm). Note that the estimate obtained by solving this problem can be seen as the best estimation of the image that we could obtain in the case when DDEs could be known perfectly (for our choice of regularization and parameters). The recovered images in this case are given in Figure 1 (2a-c) and Figure 2 (2a-c), respectively for M31 and W28. As for the original image, in both cases (1a) shows the image in linear scale, (1b) the image in log scale, and (1c) is a zoom on a subpart of the image in log scale.

In practice DDEs cannot be known perfectly. In particular, as explained in Section 2.2, we can only assume that calibration transfer has been performed and that normalized DIEs are obtained. Therefore, in practice, problem (3), with $\widetilde{r}$ given by (20), is solved to obtain an approximated estimate of $\overline{\boldsymbol{x}}$. The reconstruction obtained in this case, considering our simulation settings are given in Figure 1 (2a-c) and Figure 2 (2a-c), respectively for M31 and W28. For both cases (1a) shows the image in linear scale, (1b) the image in log scale, and (1c) is a zoom on a subpart of the image in log scale. Although, when observed in linear scale, the reconstruction seems to be of good quality, the log scale images emphasize that low amplitude coefficients are not estimated correctly. This remark is particularly true when looking to the zoomed images, where it can be observed that the structures around the bright sources have a poor reconstruction quality.

[‡]The MATLAB code is available at `https://basp-group.github.io/joint-img-dde-calibration/`.

[§]The $u-v$ tracks are simulated using the code available at `http://www.astro.umd.edu/~cychen/MATLAB/ASTR410/uvAndBeams.html`

	Image reconstruction from true DDEs	Image reconstruction from normalized DIEs	Proposed blind approach
M31	36,3	20	25,5
W28	20,6	17,2	20,6

Table 1. SNR values in dB for the images of M31 and W28 obtained by solving the imaging problem from true DDEs, the imaging problem from normalized DIEs, or the blind deconvolution problem considering DDEs.

Finally we performed simulations using our method described in Section 3, in order to jointly estimate the DDEs and the image. The reconstruction results for the images of M31 and W28 are given in Figure 1 (2a-c) and Figure 2 (2a-c), respectively. For both images, (1a) shows the image in linear scale, (1b) the image in log scale, and (1c) is a zoom on a subpart of the image in log scale. It can be observed that using the proposed method to estimate the DDEs leads to a more accurate estimation of the low amplitude coefficients of both images, and the structures around bright sources have better quality reconstructions.

In Table 1 we present the signal to noise ratio (SNR) values[¶] for each of the experiments described above. Note that, due to the ambiguity problems encountered with blind deconvolution problems, in our case we use a modified version of the SNR which is expressed as

$$\text{SNR} = 20\log_{10}\left(\frac{\|\overline{\boldsymbol{x}}\|_2}{\|\nu^\star \boldsymbol{x}^\star - \overline{\boldsymbol{x}}\|_2}\right) \quad \text{with} \quad \nu^\star = \underset{\nu>0}{\operatorname{argmin}} \ \|\nu\, \boldsymbol{x}^\star - \overline{\boldsymbol{x}}\|_2^2. \tag{21}$$

More precisely, for the image of M31 we present in the first row of the table the modified SNR values between the images given in (i) Figures 1 (2a) and (1a) in the left column (i.e. the reconstructed image obtained by considering the true DDEs), (ii) Figures 1 (3a) and (1a) (i.e. the reconstructed image obtained by considering normalized DIEs), and (iii) Figures 1 (4a) and (1a) (i.e. the reconstructed image obtained by solving the joint DDE calibration and imaging problem). Similar results are presented in the second row of the table for W28. It can be observed that for W28 the SNR value of the rescaled estimate obtained with the proposed blind approach is the same as the one obtained on the case where the true DDEs are perfectly known. Although this remark does not stand for the image of M31, it can be observed that in this case there is a difference of 5.5 dB between the SNR of the estimate obtained considering the normalized DIEs and the estimate obtained using our method.

Therefore, for both images, we can conclude that the proposed approach leads to a significant improvement in terms of quality reconstruction.

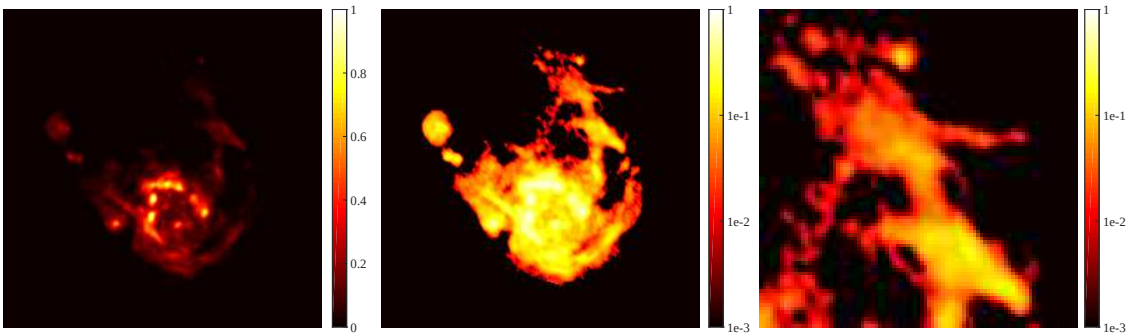
5. CONCLUSION

In this work, we have presented a non-convex optimization algorithm to jointly solve the DDE calibration and imaging problem in the context of RI. It constitutes a generalization of a previous work presented in,[10] relaxing the strong assumption made on the knowledge of the bright sources. The proposed algorithm is based on a block-coordinate forward-backward approach, using an ℓ_1 based regularization term to promote sparsity on the image and modelling the DDEs as smooth functions of the sky, i.e. spatially band-limited. Our method presents several advantages. Firstly, it benefits from convergence guarantees for both the image and the DDEs. Secondly, in contrast with DDE calibration methods developed recently, our method does not require a selection of calibrator directions since it constructs a smooth DDE screen applied to all the sources across the image. Finally, the proposed method is very general and can be easily adapted to the nature of the considered image. The good performance of the proposed method has been shown through simulations on the images of M31 and W28. In particular, the presented results have been obtained by considering that only calibration transfer has been performed, without prior knowledge on the bright sources of the images.

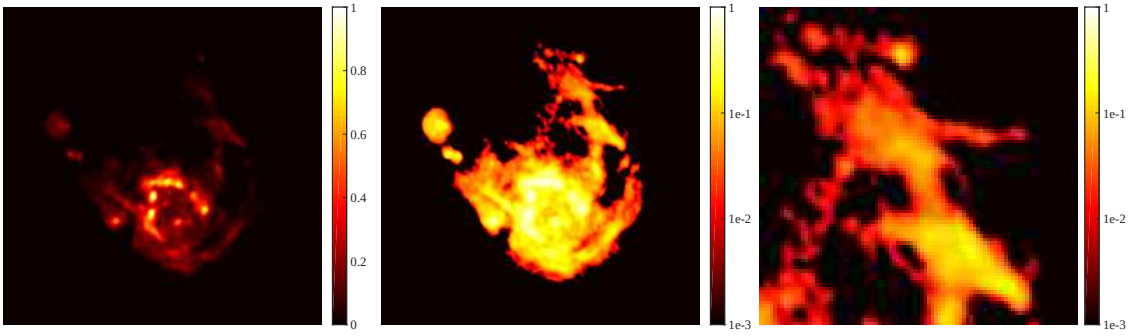
ACKNOWLEDGMENTS

This work was supported by the UK Engineering and Physical Sciences Research Council (EPSRC, grant EP/M008843/1).

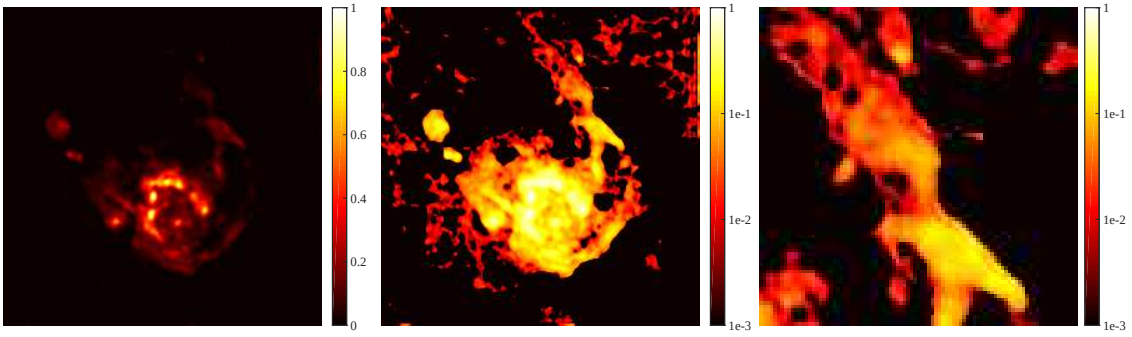
[¶]The SNR between an original image $\overline{\boldsymbol{x}}$ and its estimate $\boldsymbol{x}^\star$ is expressed as $\text{SNR} = 20\log_{10}\left(\frac{\|\overline{\boldsymbol{x}}\|_2}{\|\boldsymbol{x}^\star - \overline{\boldsymbol{x}}\|_2}\right)$.



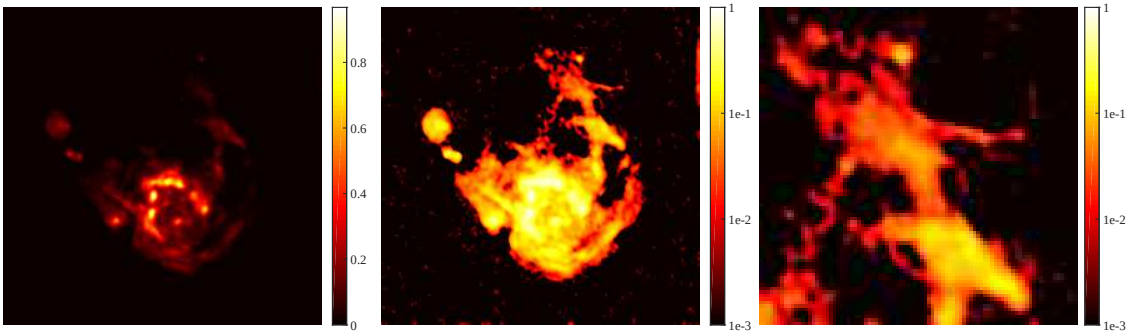
(1a) (1b) (1c)



(2a) (2b) (2c)



(3a) (3b) (3c)



(4a) (4b) (4c)

Figure 1. Row 1 shows the original image of M31 in linear scale (1a), log scale (1b) and a zoom on the top right of M31 (1c). Row 2 shows the reconstructed image of M31 obtained when solving only the imaging problem, considering the true DDEs in the reconstruction process. The reconstruction is given in linear scale (2a), log scale (2b) and a zoom on the top right of M31 in log scale (2c). Row 3 shows the reconstructed image of M31 obtained when solving only the imaging problem, considering normalized DIEs in the reconstruction process. The reconstruction is given in linear scale (3a), log scale (3b) and a zoom on the top right of M31 in log scale (3c). Row 4 shows the reconstructed image of M31 obtained when solving the joint DDE claibration and imaging problem. The reconstruction is given in linear scale (4a), log scale (4b) and a zoom on the top right of M31 in log scale (4c).

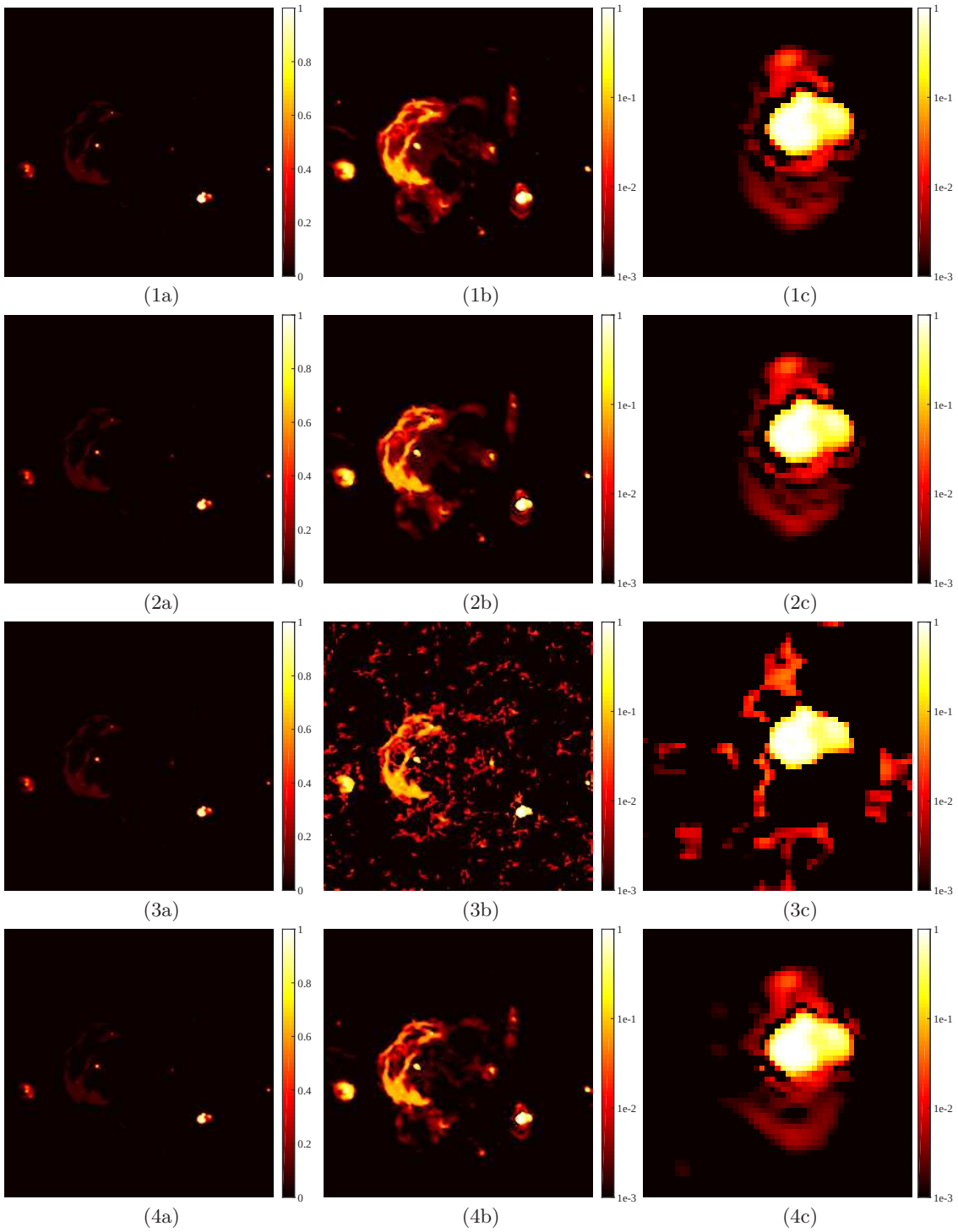


Figure 2. Row 1 shows the original image of W28 in linear scale (1a), log scale (1b) and a zoom on the top right of W28 (1c). Row 2 shows the reconstructed image of W28 obtained when solving only the imaging problem, considering the true DDEs in the reconstruction process. The reconstruction is given in linear scale (2a), log scale (2b) and a zoom on the top right of W28 in log scale (2c). Row 3 shows the reconstructed image of W28 obtained when solving only the imaging problem, considering normalized DIEs in the reconstruction process. The reconstruction is given in linear scale (3a), log scale (3b) and a zoom on the top right of W28 in log scale (3c). Row 4 shows the reconstructed image of W28 obtained when solving the joint DDE claibration and imaging problem. The reconstruction is given in linear scale (4a), log scale (4b) and a zoom on the bottom right of W28 in log scale (4c).

REFERENCES

[1] Combettes, P. L. and Pesquet, J.-C., "Proximal splitting methods in signal processing," in [*Fixed-Point Algorithms for Inverse Problems in Science and Engineering*], Bauschke, H. H., Burachik, R., Combettes, P. L., Elser, V., Luke, D. R., and Wolkowicz, H., eds., 185–212, Springer-Verlag, New York (2010).

[2] Combettes, P. L. and Wajs, V. R., "Signal recovery by proximal forward-backward splitting," *Multiscale Model. Simul.* **4**, 1168–1200 (Nov. 2005).

[3] Tseng, P., "A modified forward-backward splitting method for maximal monotone mappings," *SIAM J. Control Optim.* **38**(2), 431–446 (1998).

[4] Beck, A. and Teboulle, M., "Fast gradient-based algorithms for constrained total variation image denoising and deblurring problems," *IEEE Trans. Image Process.* **18**, 2419 –2434 (Nov. 2009).

[5] Komodakis, N. and Pesquet, J.-C., "Playing with duality: An overview of recent primal-dual approaches for solving large-scale optimization problems," *IEEE Signal Process. Mag.* **32** (Nov. 2015).

[6] Condat, L., "A primal-dual splitting method for convex optimization involving Lipschitzian, proximable and linear composite terms," *J. Optim. Theory Appl.* **158**, 460–479 (Aug. 2013).

[7] Vũ, B. C., "A splitting algorithm for dual monotone inclusions involving cocoercive operators," *Advances in Computational Mathematics* **38**(3), 667–681 (2013).

[8] Boyd, S. and Vandenberghe, L., [*Convex Optimization*], Cambridge University Press (2004).

[9] Smirnov, O., "Revisiting the radio interferometer measurement equation. ii. calibration and direction-dependent effects," *A & A* **527**, 10 (March 2015).

[10] Repetti, A., Birdi, J., Dabbech, A., and Wiaux, Y., "Non-convex optimization for self-calibration of directiondependent effects in radio interferometric imaging," *To appear in Mon. Not. R. Astron. Soc.* (2017).

[11] Uecker, M., Lai, P., Murphy, M. J., Virtue, P., Elad, M., Pauly, J. M., Vasanawala, S. S., and Lustig, M., "Espirit - an eigenvalue approach to autocalibrating parallel mri: Where sense meets grappa," *Magn. Reson. Med.* **71**, 990–1001 (2014).

[12] Högbom, J. A., "Aperture synthesis with a non-regular distribution of interferometer baselines," *A&A* **15**, 417–426 (1974).

[13] Schwarz, U. J., "Mathematical-statistical description of the iterative beam removing technique (method clean)," *A&A* **65**, 345 (1978).

[14] Thompson, A. R., Moran, J. M., and Swenson, G. W., [*Interferometry and Synthesis in Radio Astronomy*], Wiley-Interscience, New York (2001).

[15] Cornwell, T., "Multiscale CLEAN deconvolution of radio synthesis images," *IEEE J. Sel. Topics Signal Process.* **2**, 793–801 (2008).

[16] Bhatnagar, S. and Cornwell, T. J., "Scale sensitive deconvolution of interferometric images," *A&A* **426**, 747–754 (2004).

[17] Wiaux, Y., Jacques, L., Puy, G.and Scaife, A. M. M., and Vandergheynst, P., "Compressed sensing imaging techniques for radio interferometry," *Mon. Not. R. Astron. Soc.* **395**, 1733–1742 (2009).

[18] Wiaux, Y., Puy, G., Boursier, Y., and Vandergheynst, P., "Spread spectrum for imaging techniques in radio interferometry," *Mon. Not. R. Astron. Soc.* **400**(2), 1029–1038 (2009).

[19] Wenger, S., Magnor, M., Pihlström, Y., Bhatnagar, S., and Rau, U., "SparseRI: A compressed sensing framework for aperture synthesis imaging in radio astronomy," *Publications of the Astronomical Society of the Pacific* **122**(897), 1367–1374 (2010).

[20] Li, F., Cornwell, T. J., and de Hoog, F., "The application of compressive sampling to radio astronomy," *A & A* **528**(A31), 10 (2011).

[21] Carrillo, R., McEwen, J., and Wiaux, Y., "Sparsity Averaging Reweighted Analysis (SARA): a novel algorithm for radio-interferometric imaging," *Mon. Not. R. Astron. Soc.* **426**, 1223–1234 (2012).

[22] Dabbech, A., Ferrari, C., Mary, D., Slezak, E., Smirnov, O., and Kenyon, J. S., "MORESANE: MOdel REconstruction by Synthesis-ANalysis Estimators. A sparse deconvolution algorithm for radio interferometric imaging," *A & A* **576**(A7), 16 (2015).

[23] Onose, A., Carrillo, R. E., Repetti, A., McEwen, J. D., Thiran, J.-P., Pesquet, J.-C., and Wiaux, Y., "Scalable splitting algorithms for big-data interferometric imaging in the ska era," *Mon. Not. R. Astron. Soc.* **462**, 4314–4335 (2016).

[24] Mallat, S., [*A Wavelet Tour of Signal Processing*], Academic Press, Burlington, MA, 2rd ed. (2009).

[25] Rudin, L. I., Osher, S., and Fatemi, E., "Nonlinear total variation based noise removal algorithms," *Phys. D* **60**, 259–268 (1992).

[26] Gilboa, G. and Osher, S., "Nonlocal operators with applications to image processing," *Multiscale Model. Simul.* **7**(3), 1005–1028 (2008).

[27] Intema, H. T., Van der Tol, S., Cotton, W. D., Cohen, A. S., Van Bemmel, I., and Röttgering, H. J. A., "Ionospheric calibration of low frequency radio interferometric observations using the peeling scheme. i. method description and first results," *A&A* **501**(3), 1185–1205 (2009).

[28] Yatawatta, S., Zaroubi, S., Bruyn, A. G. d., Koopmans, L. V. E., and Noordam, J., "Radio interferomietric calibration using the SAGE algorithm," *Proc. 13th IEEE DSP workshop* **13**, 150–155 (2009).
[29] Kazemi, S., Yatawatta, S., Zaroubi, S., Bruyn, A. G. d., Koopmans, L. V. E., and Noordam, J., "Radio interferomietric calibration using the SAGE algorithm," *Mon. Not. R. Astron. Soc.* **414**, 1656–1666. (2011).
[30] Smirnov, O. and Tasse, C., "Radio interferometric gain calibration as a complex optimization problem," *Mon. Not. R. Astron. Soc.* **449**, 2668–2684 (2015).
[31] Tasse, C., "Nonlinear kalman filters for calibration in radio interferometry," *A & A* **556**(A127), 11 (2014).
[32] Tasse, C., "Applying wirtinger derivatives to the radio interferometry calibration problem," tech. rep. (Oct. 2014). https://arxiv.org/abs/1410.8706.
[33] van Weeren, R. J., Williams, W. L., Hardcastle, M. J., and Shimwell, T. W. e. a., "LOFAR facet calibration," *The Astrophysical Journal Supplement Series* **223**, 17 (2016).
[34] Mitchell, D. A., Greenhill, L. J., Wayth, R. B., Sault, R. J., Lonsdale, C. J., Cappallo, R. J., Morales, M. F., and Ord, S. M., "Real-time calibration of the murchison widefield array," *IEEE J. Sel. Topics Signal Process* **2**, 707–717 (Oct. 2008).
[35] Salvini, S. and Wijnholds, S. J., "Fast gain calibration in radio astronomy using alternating direction implicit methods: Analysis and applications," *A&A* **571**, A97 (2014).
[36] Bolte, J., Sabach, S., and Teboulle, M., "Proximal alternating linearized minimization for nonconvex and nonsmooth problems," *Math. Program.* **146**(1), 459–494 (2014).
[37] Frankel, P., Garrigos, G., and Peypouquet, J., "Splitting methods with variable metric for Kurdyka-Łojasiewicz functions and general convergence rates," *J. Optim. Theory Appl.* **165**, 874–900 (June 2015).
[38] Chouzenoux, E., Pesquet, J.-C., and Repetti, A., "A block coordinate variable metric forward-backward algorithm," *J. Global Optim.* **66**, 457–485 (Nov. 2016).
[39] Repetti, A., Birdi, J., and Wiaux, Y., "Joint imaging and ddes calibration for radio interferometry," 25, Biomedical and Astronomical Signal Processing (BASP) Frontiers workshop, Villars-sur-Ollon, Switzerland (29 Jan.–3 Feb. 2017).
[40] Repetti, A., Birdi, J., and Wiaux, Y., "Non-convex blind deconvolution approach for sparse radio-interferometric imaging," 130, SPARS, Lisbon, Portugal (5-8 June 2017).
[41] Wiaux, Y., Puy, G., and Vandergheynst, P., "Compressed sensing reconstruction of a string signal from interferometric observations of the cosmic microwave background," *Mon. Not. R. Astron. Soc.* **402**(4), 2626–2636 (2010).
[42] Starck, J.-L. and Murtagh, F., "Image restoration with noise suppression using the wavelet transform," *A & A* **288**, 342–348 (Aug. 1994).
[43] Garsden, H., Girard, J. N., Starck, J. L., Corbel, S., Tasse, C., Woiselle, A., McKean, J. P., van Amesfoort, A. S., Anderson, J., Avruch, I. M., Beck, R., Bentum, M. J., Best, P., Breitling, F., Broderick, J., Brüggen, M., Butcher, H. R., Ciardi, B., de Gasperin, F., de Geus, E., de Vos, M., Duscha, S., Eislöffel, J., Engels, D., Falcke, H., Fallows, R. A., Fender, R., Ferrari, C., Frieswijk, W., Garrett, M. A., Grießmeier, J., Gunst, A. W., Hassall, T. E., Heald, G., Hoeft, M., Hörandel, J., van der Horst, A., Juette, E., Karastergiou, A., Kondratiev, V. I., Kramer, M., Kuniyoshi, M., Kuper, G., Mann, G., Markoff, S., McFadden, R., McKay-Bukowski, D., Mulcahy, D. D., Munk, H., Norden, M. J., Orru, E., Paas, H., Pandey-Pommier, M., Pandey, V. N., Pietka, G., Pizzo, R., Polatidis, A. G., Renting, A., Röttgering, H., Rowlinson, A., Schwarz, D., Sluman, J., Smirnov, O., Stappers, B. W., Steinmetz, M., Stewart, A., Swinbank, J., Tagger, M., Tang, Y., Tasse, C., Thoudam, S., Toribio, C., Vermeulen, R., Vocks, C., van Weeren, R. J., Wijnholds, S. J., Wise, M. W., Wucknitz, O., Yatawatta, S., Zarka, P., and Zensus, A., "LOFAR sparse image reconstruction," *A&A* **575**, A90 (Mar. 2015).
[44] Vannier, M., Mary, D., Millour, F., Petrov, R. G., Bourguignon, S., and Theys, C., "Spectral regularization and sparse representation bases for interferometric imaging," *Proc. SPIE* **7734**, 77342J–77342J–15 (July 2010).
[45] Rockafellar, R. T., [*Convex Analysis*], Princeton University Press (1970).
[46] Bauschke, H. H. and Combettes, P. L., [*Convex Analysis and Monotone Operator Theory in Hilbert Spaces*], Springer, New York (2011).
[47] Rockafellar, R. T. and Wets, R. J.-B., [*Variational Analysis*], Springer-Verlag, 1st ed. (1997).
[48] Mordukhovich, B. S., [*Variational Analysis and Generalized Differentiation. Vol. I: Basic theory*], vol. 330 of *Series of Comprehensive Studies in Mathematics*, Springer-Verlag, Berlin-Heidelberg (2006).
[49] Moreau, J. J., "Proximité et dualité dans un espace hilbertien," *Bull. Soc. Math. France* **93**, 273–299 (1965).
[50] Luo, Z. Q. and Tseng, P., "On the convergence of the coordinate descent method for convex differentiable minimization," *J. Optim. Theory Appl.* **72**(1), 7–35 (1992).
[51] Bolte, J., Combettes, P. L., and Pesquet, J.-C., "Alternating proximal algorithm for blind image recovery," in [*Int. Conf. Image Process. (ICIP 2010)*], **4**, 1673–1676 (26-29 Sep. 2010).
[52] Xu, Y. and Yin, W., "A block coordinate descent method for regularized multiconvex optimization with applications to nonnegative tensor factorization and completion," **6**(3), 1758–1789 (2013).